Supporting Online Material for:

PhIP-Seq: a Method for Autoantigen Discovery

H. Benjamin Larman[1,2,3], Zhenming Zhao[3,4], Uri Laserson[1,5,6], Mamie Z. Li[3], Alberto Ciccia[3], M. Angelica Martinez Gakidis[3], George M. Church[6], Santosh Kesari[7], Emily M. LeProust[8], Nicole L. Solimini[3*] & Stephen J. Elledge[3*]

[1]Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA

[2]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

[3]Department of Genetics, Harvard University Medical School, and Division of Genetics, Howard Hughes Medical Institute, Brigham and Women's Hospital, Boston, MA, USA

[4]Present address: Biogen Idec, Cambridge, Massachusetts, USA

[5]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

[6]Department of Genetics, Harvard University Medical School, Boston, MA, USA

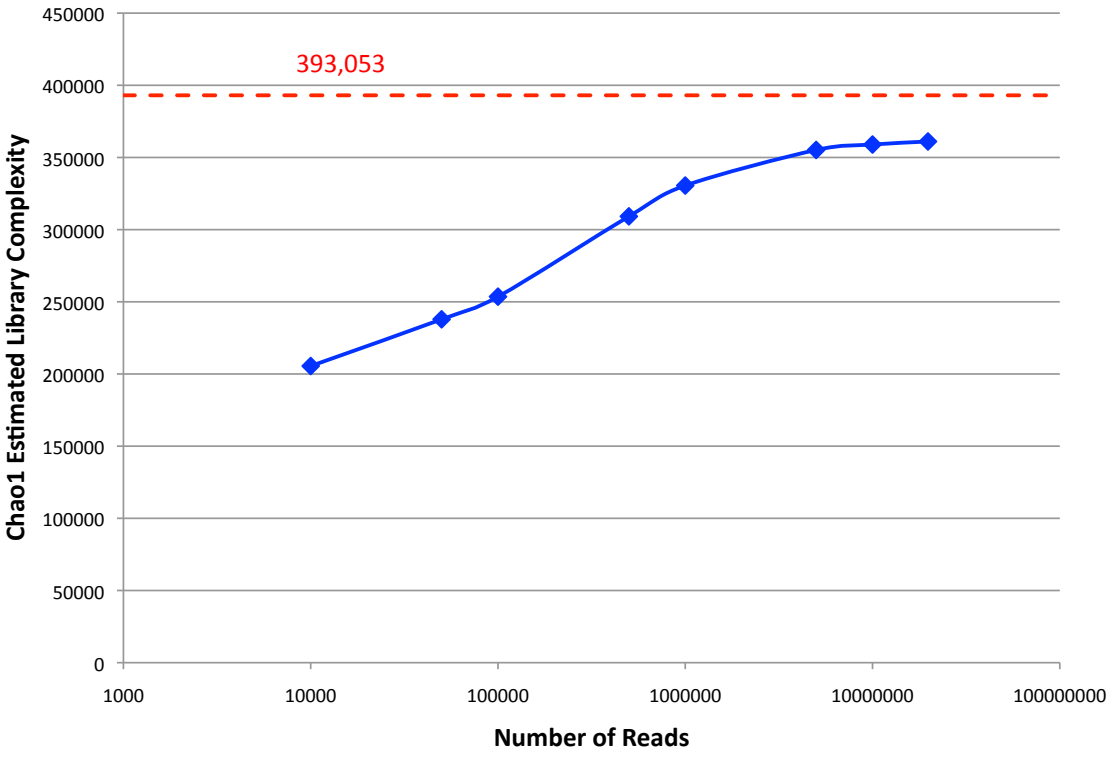
[7]Division of Neuro-Oncology, Department of Neurosciences, U.C. San Diego, Moores Cancer Center, La Jolla, CA, USA

[8]Agilent Technologies, Genomics, Santa Clara, CA, USA

*Correspondence to:
nsolimini@rics.bwh.harvard.edu, selledge@genetics.med.harvard.edu

This file contains:

Supplementary Figure 1.



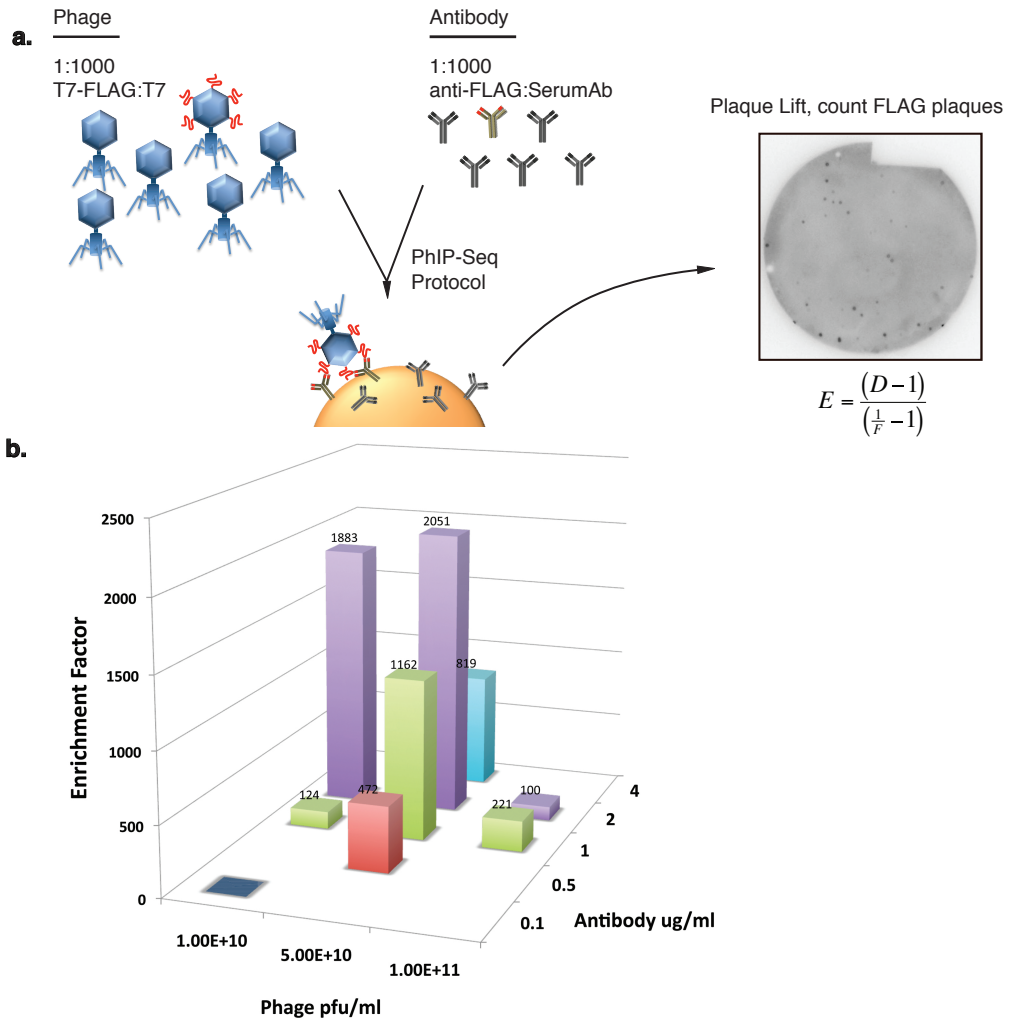
Supplementary Figure 1. The effect of sequencing depth on estimated library complexity.

Chao1 estimates of library complexity given by

$$S_{Chao1} = S_{obs} + \frac{n_1^2}{2n_2}$$

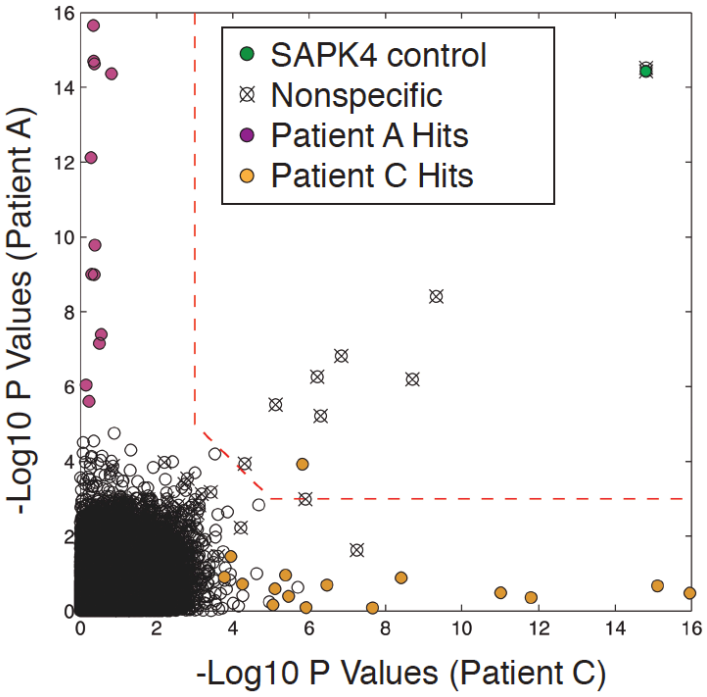
are shown as a function of simulated T7-Pep library sampling. S_{Chao1} is the estimated complexity, where S_{obs} is the observed library complexity, n_1 is the number of library members observed once, and n_2 is the number of library members observed twice. For the data points shown, S_{obs}, n_1, and n_2 were simulated by randomly sampling the actual sequencing data "Number of Reads" times without replacement. S_{Chao1} was then calculated as above. The sequencing depth actually achieved, ~20 million reads, appears to be near saturating with respect to Chao1 estimate of the library complexity, at 361,070 library members (or ~91.8% of the 393,053 resolvable clones).

Supplementary Figure 2.



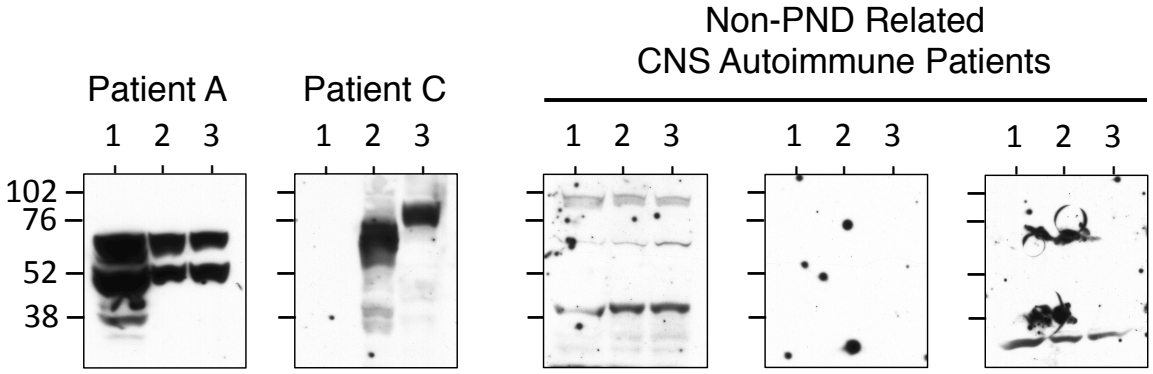
Supplementary Figure 2. Optimization of PhIP-Seq target enrichment
(a) A FLAG-expressing T7 phage (depicted with red peptide) was diluted at 1:1,000 into native, non-FLAG-expressing T7 phage to mimic a target peptide within the T7-Pep library. An anti-FLAG monoclonal antibody (M2, Sigma-Aldrich; shown with red variable region) was diluted 1:1,000 into human serum antibodies (shown with black variable region) to mimic a rare autoantibody within a patient's antibody repertoire. After performing the IP, plaque lift analysis for FLAG expression was performed to determine enrichment using the equation shown (*E* = enrichment; *D* = dilution factor = 1,000; *F* = fraction of FLAG expressing clones on plaque lift). Enrichment was optimized with respect to type of beads, number of washes, order of antibody-phage/antibody-bead complex formation, and relative concentrations of phage and antibody. **(b)** Enrichment factor was found to depend on the relative concentrations of phage and antibody during complex formation. We thus varied these parameters independently and found an optimum at about 5×10^{10} pfu/ml phage and 2 μg/ml total antibody.

Supplementary Figure 3.



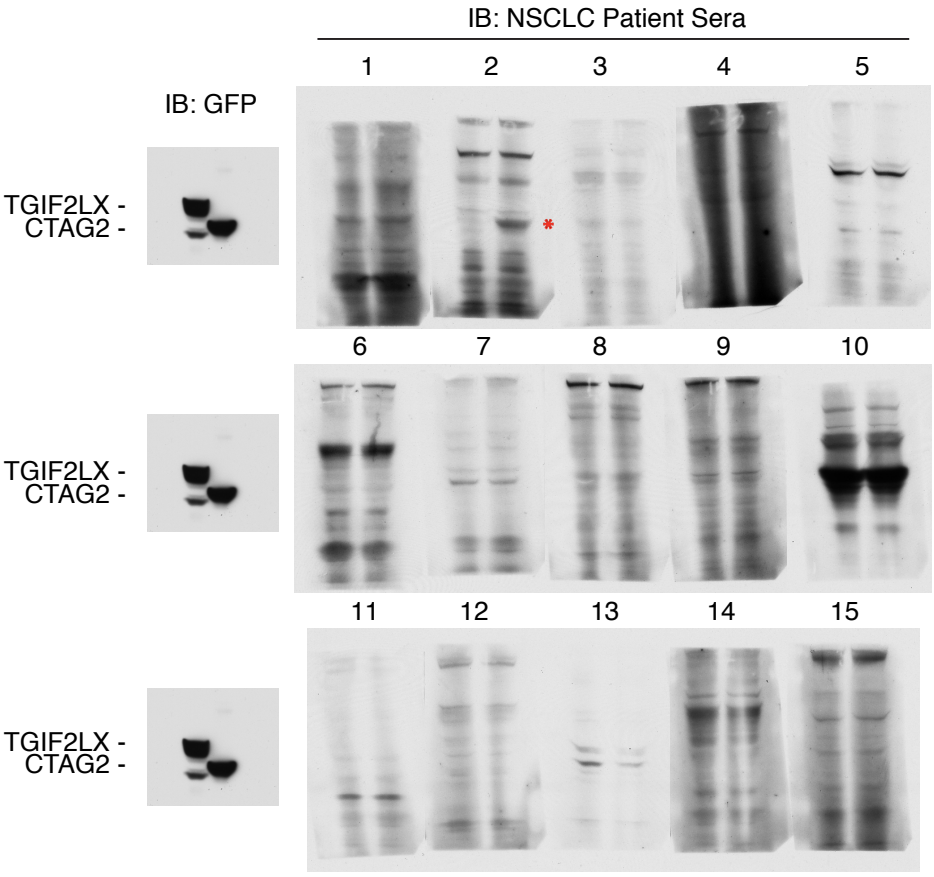
Supplementary Figure 3. Comparison of PhIP-Seq experiments on different patients.
Scatter plot as in Fig. 2d from text, but comparing clone enrichment p-values from two different patients: Patient A (y-axis) versus Patient C (x-axis). Both experiments included the SAPK4 spike-in antibody. X'ed circles were enriched by beads and SAPK4 antibody alone (no patient antibody in IP). Filled purple and orange circles are the Patient A- and Patient C-specific positives given in Table 2 from the text.

Supplementary Figure 4.



Supplementary Figure 4. TGIF2LX, TRIM9 and TRIM67 autoreactivity is not present nonspecifically in CSF.
Western blotting with CSF from Patients A and C, as well as three patients with non-PND related CNS autoimmune syndromes. In each blot, lanes 1, 2, and 3 were loaded with lysate from 293T cells overexpressing either TGIF2LX-GFP, FLAG-TRIM9, or FLAG-TRIM67, respectively.

Supplementary Figure 5.



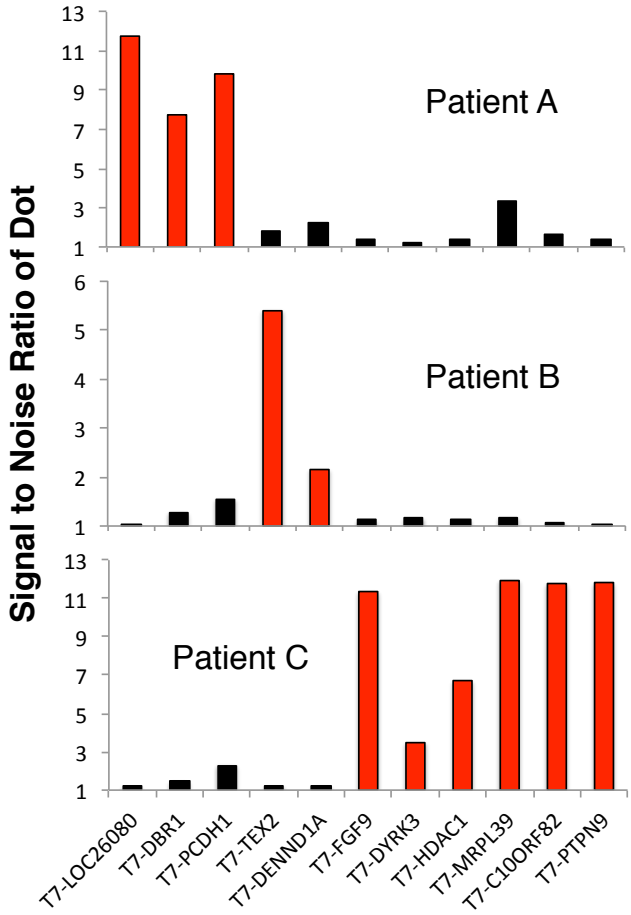
Supplementary Figure 5. Immunoblots for TGIF2LX and CTAG2 reactivity in the serum of NSCLC patients without PND.
Sera from fifteen non-small cell lung cancer (NSCLC) patients were used to blot SDS-PAGE separated 293T cell lysate overexpressing either TGIF2LX (left lane) or CTAG2 (right lane), fused with C-terminal GFP. Staining for GFP (left blot) demonstrates overexpression of TGIF2LX and CTAG2 at the expected weights. Only patient 2 was found to have anti-CTAG2 serum antibodies (marked by *). No patients were found to have anti-TGIF2LX serum antibodies.

Supplementary Figure 6.

Gene	T7-Pep Clone	Peptides Enriched by Patient C	-Log10 P Value
TRIM9	NP_443210.1_2	LD---------L]	0.4
	NP_443210.1_3	LD---------LDKMSLYSEADSGYGSYGGFASAPTTPCQK]	0.9
	NP_443210.1_4	[PTTPCQKSPNGVRVFPPAMPPPATHLSPALAPVPR----	1.6
	NP_443210.1_5	[LAPVPR----	0.7
TRIM67	NP_001004342.2_4	[LGGGAGGGGDHADKLSLYSETDSGYGSY--------TPSLKSPN]	15.7
	NP_001004342.2_5	[PSLKSPNGVRVLPMVPAPPGSSAAAARGAACSSLSS]	5.3
		*. **:*****:******* **. ********:* . .**.: :.* .. .	
TRIM9	NP_443210.1_6	[PKNRVLEGVIDRYQQSK---------AAALKCQLCEKAP-KEATVM]	15.6
TRIM67	NP_001004342.2_6	QRNRL]	1.5
	NP_001004342.2_7	QRNRLLEAIVQRYQQGRGAVPGTSAAAAVAICQL]	0.9
	NP_001004342.2_8	[AVAICQLCDRTPPEPAATL	0.2
		:**:**.:::****.: **. ****:::* : *:.:	
TRIM9	NP_443210.1_12	[CDALIDALNRRKAQLLARVNKEHEHKLKVVRDQISH]	15.2
TRIM67	NP_001004342.2_14	CDALVDALTRQKAKLLTKVT]	0.5
	NP_001004342.2_15	[KLLTKVTKEREHKLKMVWDQINH	0.6
		****:***.*:**:**::*.**:*****: ****.*	
TRIM9	NP_443210.1_19	[AFNKTGVSPYSKTLVLQTSEGKALQQYPS----------ERELRGI]	4.1
TRIM67	NP_001004342.2_21	AFNSSGVGPYSKTVVLQTSDVAWFTFDPNS]	0.4
	NP_001004342.2_22	[FTFDPNSGHRDIILSNDNQTATCSSYDDRVVLGT	0.5
		.:**.**:*****: : *. :* : *	

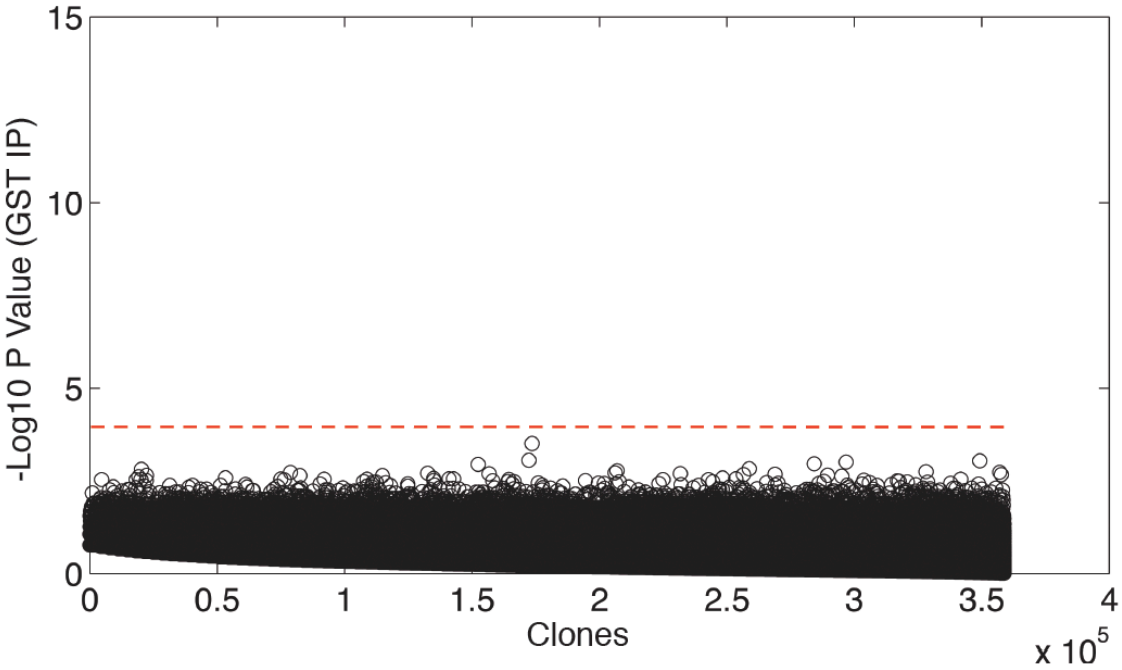
Supplementary Figure 6. Alignment among enriched peptides from TRIM9 and TRIM67.
Significantly enriched peptides (in red) from TRIM9 and TRIM67 shown with corresponding ClustalW-aligned peptides from the homologous protein (in black). Boundaries of phage-displayed peptides are denoted with brackets. Peptides are shown next to their –Log10 p-value of enrichment.

Supplementary Figure 7.



Supplementary Figure 7. Quantification of T7 Candidate Dot Blots.
The dot blots in Fig. 3g were analyzed to determine the signal-to-noise ratio arising from each T7 candidate clone immunoblotted with each of the patients' spinal fluid. The data from the candidates expected to react with a given patient's antibodies are shown in red, whereas that data from the candidates that are expected not to react with a given patient's antibodies are shown in black.

Supplementary Figure 8



Supplementary Figure 8. PhIP-Seq –Log10 p-values for T7-Pep enrichment by GST alone.
GST coated glutathione magnetic beads were used to precipitate phage from the T7-Pep library. Illumina sequencing data was analyzed using the generalized Poisson method. No library members were significantly enriched by GST alone ($P < 10^{-4}$).

Supplementary Table 1.

Pool	Plaques analyzed	Plaques with multiple inserts	% Multiple inserts	Plaques with vector religation	% Vector Religation
T7-Pep pool 1	45	1	2.2	1	2.2
T7-Pep pool 2	39	3	7.7	0	0.0
T7-Pep pool 3	39	1	2.6	0	0.0
T7-Pep pool 4	38	3	7.9	0	0.0
T7-Pep pool 5	38	2	5.3	0	0.0
T7-Pep pool 6	39	0	0.0	0	0.0
T7-Pep pool 7	31	1	3.2	0	0.0
T7-Pep pool 8	62	3	4.8	1	1.6
T7-Pep pool 9	54	0	0.0	0	0.0
T7-Pep pool 10	31	1	3.2	0	0.0
T7-Pep pool 11	62	3	4.8	1	1.6
T7-Pep pool 12	69	1	1.4	4	5.8
T7-Pep pool 13	31	0	0.0	0	0.0
T7-Pep pool 14	31	1	3.2	0	0.0
T7-Pep pool 15	31	1	3.2	1	3.2
T7-Pep pool 16	31	0	0.0	1	3.2
T7-Pep pool 17	30	1	3.3	0	0.0
T7-Pep pool 18	30	1	3.3	0	0.0
T7-Pep pool 19	31	1	3.2	0	0.0
T7-NPep pool 1	46	3	6.5	1	2.2
T7-CPep pool 1	47	2	4.3	0	0.0
T7-NPep pool 2	48	0	0.0	3	6.3
T7-CPep pool 2	44	1	2.3	1	2.3
Total	**947**	**30**	**3.2**	**14**	**1.5**

Supplementary Table 1. Subpool analysis of multiple insertions and vector re-ligation after cloning of the T7-Pep, T7-NPep, and T7-CPep libraries. Phage plaques from each subpool were randomly selected and PCR analyzed to examine the frequency of multiple insertions and vector religations present within each pool.

Supplementary Table 2

Pool	FLAG-positive plaques	T7 tail fiber positive plaques	% in-frame phage
T7-Pep pool 2	44	69	64%
T7-Pep pool 3	61	94	65%
T7-Pep pool 4	43	64	67%
T7-Pep pool 5	48	70	69%
Total	196	297	66%

Supplementary Table 2. Subpool analysis of FLAG expression after cloning of T7-Pep. Plaque lifts from four subpools were analyzed by immunoblotting using FLAG and T7 tail fiber antibodies to measure in-frame and total plaques, respectively. Plaques staining positive were counted and a percentage of in-frame, FLAG-expressing phage was determined. The vast majority of frameshifting mutations present in the phage inserts is due to errors in DNA chemical synthesis on the releasable DNA microarrays. In parallel oligonucleotide synthesis, sequence integrity can be compromised by depurination side reactions, inefficient nucleoside coupling, and reversible 5'-hydroxyl deprotection reactions, leading to mutations of the desired oligonucleotide. Great strides have been made to improve the fidelity of this technology[33], resulting in a per-cycle yield of 99.5%. However, since oligonucleotide synthesis for our libraries required as many as 140 cycles, even very small per-cycle losses can result in a significant loss of fidelity.

Supplementary Table 3.

Rank	T7-Pep Clone	Peptide	Log P GST	Log P GST-RPA2	Gene Symbol
1	**NP_054859.2_1**	**MSLPLTEEQRKKIEENRQKALARRAEKLLAEQHQRT**	**0.29**	**14.61**	**SMARCAL1**
2	NP_055877.3_31	TPPSMSAALPFPAGGLGMPPSLPPPPLQPPSLPLSM	1.09	6.60	PPRC1
3	NP_006360.3_18	TLSYNGLGSNIFRLLDSLRALSGQAGCRLRALHLSD	2.23	6.59	LRRC41
4	NP_060903.2_28	AVLQQNPSVLEPAAVGGEAASKPAGSMKPACPASTS	0.07	5.95	KDM3A
5	XP_372311.2_13	LTLYDGPNVSSPSYGPYCRGDTSIAPFVASSNQVFI	1.65	5.90	LOC389958
6	NP_060876.2_13	LTPVTTSTVLSSPSGFNPSGTVSQETFPSGETTISS	1.87	5.68	MUC4
7	XP_372592.2_4	AALIHVPPLSRGLPASLLGRALRVIIQEMLEEVGKP	0.28	5.39	PGPEP1L
8	NP_057131.1_2	ITAEEMYDIFGKYGPIRQIRVGNTPETRGTAYVVYE	0.47	5.26	SF3B14
9	**NP_003353.1_3**	**AEQLDRIQRNKAAALLRLAARNVPVGFGESWKKHLS**	**0.20**	**5.23**	**UNG2**
10	NP_443728.2_10	IRPMDDDLLKLLLPLMLQYSDEFVQSAYLSRRLAYF	1.32	5.06	MED12L
11	NP_004981.2_2	ISTVGPEDCVVPFLTRPKVPVLQLDSGNYLFSTSAI	1.60	5.00	MARS
12	NP_078997.2_120	TTSTSQSAASSNNTYPHLSCFSMKSWPNILFQASAR	0.22	4.96	ZFHX4
13	NP_004697.2_27	ITETAGSLKVPAPASRPKPRPSPSSTREPLLSSSEN	2.14	4.96	ARHGEF1
14	NP_003425.2_23	SHLSRHRKTTSVHHRLPVQPDPEPCAGQPSDSLYSL	1.28	4.81	ZNF133
15	NP_996882.1_34	LDRFKNRLKDYPQYCQHLASISHFMQFPHHLQEYIE	0.54	4.80	CNOT1
16	NP_783324.1_3	PHPSALSSVPIQANALDVSELPTQPVYSSPRRLNCA	2.11	4.71	RAB3IP
17	NP_000341.1_5	PESQHLGRIWTELHILSQFMDTLRTHPERIAGRGIR	0.02	4.70	ABCA4
18	NP_689896.1_1	MNRKWEAKLKQIEERASHYERKPLSSVYRPRLSKPE	0.78	4.68	CCDC111
19	XP_095991.7_15	IKTRDICNQLQQPGFPVTVTVESPSSSEVEEVDDSS	1.04	4.65	CEP78
20	NP_741996.1_43	ITNGLAMKNNEISVIQNGGIPQLPVSLGGSALPPLG	1.60	4.60	SALL3
21	NP_997191.1_49	SVYGWATLVSERSKNGMQRILIPFIPAFYINQSELV	0.86	4.48	NUP210L
22	NP_775902.2_9	IHSGERPYECSECGKLFMWSSTLITHQRVHTGKRPY	1.31	4.41	ZNF547
23	XP_496363.1_6	PVRRGYWGNKIGKPHTVPCKVTGRCGSALVHLIPVP	1.66	4.34	RNF11
24	NP_001026.1_92	LSRKLFWGIFDALSQKKYEQELFKLALPCLSAVAGA	0.05	4.29	RYR2
25	NP_006359.3_11	THQWLDGSDCVLQAPGNTSCLLHYMPQAPSAEPPLE	0.78	4.24	CREB3
26	NP_002146.2_6	ITVPAYFNDSQRQATKDAGAIAGLNVLRIINEPTAA	1.37	4.23	HSPA6
27	NP_065987.1_4	ITPTRELAIQIDEVLSHFTKHFPEFSQILWIGGRNP	2.50	4.18	DDX55
28	NP_079279.2_13	SHHDTAVLITRYDICSSKEKCNMLGLSYLGTICDPL	0.24	4.17	ADAMTS20
29	NP_055987.1_47	PKGEPTRRGRGGTFRRGGRDPGGRPSRPSTLRRPAY	1.62	4.13	BAT2L2
30	NP_005712.1_2	IIPSCIAIKESAKVGDQAQRRVMKGVDDLDFFIGDE	1.82	4.13	ACTR3
31	NP_079165.3_2	SPPSQLFSSVTSWKKRFFILSKAGEKSFSLSYYKDH	1.75	4.13	C10ORF81
32	NP_006609.2_26	PPYKYKLRYRYTLDDLYPMMNALKLRAESYNEWALN	0.63	4.08	LOC100133760
33	NP_149163.2_41	INLTIRGHEVVGIVGRTGSGKSSLGMALFRLVEPMA	1.01	4.04	ABCC11
34	NP_001004750.1_11	IHFLFPPFMNPFIYSIKTKQIQSGILRLFSLPHSRA	0.79	4.03	OR51B6

Supplementary Table 3. Candidate RPA2 interacting proteins.
PhIP-Seq was performed using GST-RPA2 as bait, and enrichment scores (–Log10 p-values estimated by the generalized Poisson method) were compared to enrichment on GST alone.

Supplementary Methods

In Silico design of T7-Pep, T7-CPep, and T7-NPep

We first downloaded all human protein and cDNA sequences available from the RefSeq database at build 35.1 of the human genome. Accession numbers between a protein and its cDNA were matched, and the paired sequences were used to construct the library. All the ATG start codons in the cDNAs were compared to the corresponding protein sequences until the correct ORF sequence was found. 72 nt fragments were then separated and overlapped with adjoining sequences by 21 nt (7 aa). Each DNA fragment was then scanned for the 8 relatively rare codons in *E. coli* (CTA, ATA, CCC, CGA, CGG, AGA, AGG, GGA), and they were replaced by more abundant, synonymous codons (selected randomly if there was more than one replacement available). After that, each DNA fragment was rescanned for the four restriction sites (EcoRI, XhoI, BseRI, MmeI), and they were eliminated by replacement of one codon with a different, abundant, synonymous codon. Sequences were scanned iteratively to ensure the final ORF fragments were free of both rare codons and restriction sites. Finally, common primer sequences were added.

Cloning of T7-Pep

The proteome-wide library (19 pools of 22,000 synthetic oligos per pool) and N/C-terminal libraries (2 pools each of 18,000 synthetic oligos per pool) were PCR-amplified as 23 independent pools with common primer sequences using the following conditions: 250 mM dNTPs, 2.5 mM $MgCl_2$, 0.5 μM each primer, 1 μl Taq polymerase and ~350 ng oligo DNA per 50 μl reaction. The thermal profile was 1. 95°C 30 sec, 2. 94°C 35 sec, 3. 50°C 35 sec, 4. 72°C 30 sec, 5. go to step 2 3x, 6. 72°C 5 min, 7. 95°C 30 sec, 8. 94°C 35 sec, 9. 70°C 35 sec, 10. 72°C 30 sec, 11. go to step 8, 29x, 12. 72°C 5 min, 13. 4°C forever.

The PCR product was then digested and cloned into the EcoRI/SalI sites of the T7FNS2 vector with an average representation of at least 100 copies of each peptide maintained during each cloning step. The T7FNS2 vector is a derivative of the T7Select 10-3b vector (Novagen), which is a lytic, mid-copy phage display system, and displays between 5-15 copies as C-terminal fusions with the T7 capsid protein. We modified the T7Select 10-3b vector to generate T7FNS2 by inserting a sequence encoding a FLAG epitope in the NotI and XhoI sites to generate an in-frame FLAG C-terminal fusion with the inserted peptide. Cloning of the synthetic peptide libraries into the T7FNS2 vector results in a C-terminal fusion of the ORF fragments with the T7 10B capsid protein, followed by a C-terminal FLAG epitope tag and stop codon (except for those in T7-CPep, which retain the native stop codons).

Patient Samples

Spinal fluid was removed from consenting patients suspected of having a paraneoplastic neurological syndrome using standard clinical procedures and according to approved IRB protocols. CSF was aliquoted and kept at –80°C until used, and freeze-thawing was avoided as much as possible after that. Neurological evaluations were performed by a board-certified neurologist. Serum samples from patients with confirmed non-small cell lung cancer were purchased from Bioserve (Beltsville, MD).

Detailed PhIP-Seq Protocol

Materials Required:

Primers:

The following are the multiplex barcode-introducing forward primers. Red is the common P5 sequence for Illumina sequencing. Underlined is where the sequencing primer anneals. Blue is the 3 nt barcode.

HsORF-FL-mmBC1-F
AATGATACGGCGACCACCGAAGGTGTGATGCTCGGGGATCCAGGAATTCCACT**GCGC**
HsORF-FL-mmBC2-F
AATGATACGGCGACCACCGAAGGTGTGATGCTCGGGGATCCAGGAATTCCGCC**GCGC**
HsORF-FL-mmBC3-F
AATGATACGGCGACCACCGAAGGTGTGATGCTCGGGGATCCAGGAATTCCCCT**GCGC**
HsORF-FL-mmBC4-F
AATGATACGGCGACCACCGAAGGTGTGATGCTCGGGGATCCAGGAATTCCTCT**GCGC**
HsORF-FL-mmBC5-F
AATGATACGGCGACCACCGAAGGTGTGATGCTCGGGGATCCAGGAATTCCGAT**GCGC**
HsORF-FL-mmBC6-F
AATGATACGGCGACCACCGAAGGTGTGATGCTCGGGGATCCAGGAATTCCGGT**GCGC**
HsORF-FL-mmBC7-F
AATGATACGGCGACCACCGAAGGTGTGATGCTCGGGGATCCAGGAATTCCGTT**GCGC**
HsORF-FL-mmBC8-F
AATGATACGGCGACCACCGAAGGTGTGATGCTCGGGGATCCAGGAATTCCGCG**GCGC**

P7-T7Down (this is the common reverse primer):
CAAGCAGAAGACGGCATACGAC ACTG AACCCCTCAAGACCCGTTTA

mmBC-FL_seq_prim (for sequencing the barcode and the library insert at P5 in forward direction):
AGGTGTGATGCTCGGGGATCCAGGAATTCC

IP Wash Buffer: 150 mM NaCl, 50 mM Tris-HCl, 0.1% NP-40 (pH 7.5)

Protocol:

Block 1.5 ml tubes (including under cap) with 3% fraction V BSA in TBST overnight at 4°C rotating. Positive control SAPK4 C-19 antibody (Santa Cruz, sc-7585) is added (2 ng/ml final concentration; 1/1,000 of patient antibody) to phage stock ($5x10^{10}$ pfu T7-Pep/ml final concentration) and mixed before being added to patient antibody (2 μg/ml final concentration). Bring each IP reaction to final volume of 1 ml using M9LB (Novagen).
Note: Replicas should be independent after this point (i.e. there should be two IP reactions as above for each sample performed in replicate).

Rotate tubes at 4°C for 24 h. Add 40 μl slurry of 1:1 mix of Protein A coated magnetic Dynabeads and Protein G coated magnetic Dynabeads (Invitrogen, 100.02D and 100.04.D) to each tube. Rotate for 4 more hours at 4°C. Wash beads 6 times in 500 μl IP wash buffer by pipetting up and down 8 times per wash. Change tubes after every second wash. Remove as much wash buffer as possible and resuspend beads in 30 μl H2O. Heat at 90°C for 10 min to denature phage and release DNA. Transfer H2O + DNA to new tubes for PCR. Prepare 50 μl PCR reactions with TaKaRa HS Ex polymerase (TAKARA BIO Inc.; RR006A), utilizing the entire 30 μl of H2O + DNA from the IP:

Takara HS Ex			
rxn vol	50		PCR
H2O	9.5		1. 98C, 10s
10x TaKaRa Buffer	5		2. 56C, 15s
dNTP	4		3. 72C, 25s
P7-BC-T7Down (100uM)	0.5		4. Go to 1 39x
P5-mmBCn-F (100uM)	0.5		5. 72C, 7 min
TaKaRa Ex HS	0.5		
Template	30		

(The number of cycles can optionally be increased to 45.)

Gel purify PCR products individually. Measure concentration and then mix 500 ng of each barcoded sample and perform Illumina sequencing on final material, using mmBC-FL_seq_prim as sequencing primer.

The first seven nt calls will arise from the DNA barcode, and should be used to parse the data by sample. Remaining sequence should be aligned against the reference file, which is available for download upon request. The reference sequences should be truncated to the length of the reads and alignment should be constrained to the appropriate orientation. Perform data analysis of enrichment significance using Matlab scripts provided as .m files. The target SAPK4 clone (NP_002745.1_13) should have –Log10 p-value of ~15. Set cutoffs for enrichments of clones to be considered positives. A list of clones enriched by beads alone can be provided upon request.

RPA2-peptide interaction screen

Full-length, sequence-verified RPA2 was recombined from an available entry vector into pDEST-15 for inducible expression in *E. coli* as an N-terminal GST-fusion protein. A pDEST-15 clone expressing GST alone was used as a negative control. Protein expression was induced with 0.1 μM IPTG for 5 hours at 30°C. Protein lysate from 50 ml of bacterial culture was prepared in 1.5 ml of lysis buffer (50 mM tris pH 7.5, 500 mM NaCl, 10% glycerol, 1% triton, 10 mg/ml lysozyme) and sonicated before removing insoluble material by centrifugation. 40 μl of MagneGST Glutathione beads (Promega, V8611) were incubated in 1 ml of undiluted bacterial lysate for 2 hours. Beads were then washed 3 times with PBS. 1 ml of M9LB containing $5x10^{10}$ pfu of T7-Pep was then used to resuspend the beads (now coated with GST or GST-RPA2). The mix was rotated 24 hours at 4°C. At this point the beads were washed 6 times in 500 μl IP wash buffer, and the remaining protocol for PhIP-Seq given above was followed precisely.

Estimation of general Poisson model parameters and regressions

We assessed several distribution families for their ability to appropriately model the PhIP-Seq enrichment data, and found the two-parameter generalized Poisson distribution to be the best:

$$pmf(x) = \theta(\theta + x\lambda)^{x-1} e^{-\theta - x\lambda} / x!$$

For each value of input read number that had at least 50 corresponding clones, we used the following maximum likelihood estimators to calculate the values of lambda (λ) and theta (θ) for the corresponding distribution of n IP reads (x_i) (Consul and Shoukri, Communications in Statistics - Theory and Methods 13, 1533-1547 (1984)).

$$\sum_{i=1}^{n} \frac{x_i(1-x_i)}{X + (x_i - X)\lambda} - nX = 0 \quad \text{where} \quad X = \sum_{i=1}^{n} \frac{x_i}{n} \quad \text{and} \quad \theta = X(1-\lambda)$$

Upon calculation of λ across all the input read numbers, we found it to be approximately constant. For each experiment, we thus regressed this parameter to be equal to the mean of all calculated λ's (Fig. 2c). Calculation of θ's for all input values revealed the near linearity of this parameter, and so we linearly regressed this parameter prior to calculating the p-values.

Western blot validation of candidate autoantigens

We utilized the ORFeome collection of full-length proteins, which was generated by PCR and Gateway recombinational cloning (Lamesch et al., Genomics. 2007, 89(3):307-15), as a source for testing autoantigen candidates by immunoblot. Entry vectors were recombined into the appropriate mammalian expression

vector (CMV promoter driving ORF expression with either C-terminal GFP fusion or N-terminal FLAG epitope tag) and miniprepped for transient transfection.

293T cells were plated 24 hours before transfection at a density of 0.8 million cells per well of a 6-well plate and grown in DMEM containing10% FBS. TransIT-293T transfection reagent (Mirus, MIR 2700) was mixed with 2 μg expression plasmid per well, and added to the cells. After 24 hours, cells were harvested in 200 μl standard 1x RIPA-based laemmli/DTT sample buffer with Complete protease inhibitor cocktail (Roche) and sonicated for 30 seconds. Insoluble material was removed by centrifugation. 2-20 μl of lysate was run on 4-20 Bis-Tris polyacrylamide gels and transferred onto nitrocellulose using the iBlot system (Invitrogen). Membranes were blocked 1 hr in 5% milk and then stained with either patient CSF (1:250 to 1:1,000) or the appropriate primary anti-GFP (JL-8 monoclonal antibody; Clontech, 632381) or anti-FLAG (M2 monoclonal antibody; Sigma-Aldrich, F9291) antibody in 2.5% milk, TBST. Human antibody from CSF was detected with 1:3,000 peroxidase-conjugated goat affinity purified anti-Human IGG (whole molecule) secondary antibody (MP Biomedicals, 55252) in 2.5% milk, TBST.

For IP-western blotting, cell lysate was harvested in standard RIPA buffer with Complete protease inhibitor cocktail and sonicated for 30 seconds. Insoluble material was removed by centrifugation. 150 μl of lysate was mixed with 1 μg of patient antibodies and rotated overnight at 4°C. A 40 μl slurry of 1:1 mix of Protein A coated magnetic Dynabeads and Protein G coated magnetic Dynabeads was added to each tube. Tubes were rotated 4 hours at 4°C. Beads were washed 3 times in 500 μl RIPA buffer, and then harvested in 25 μl of laemmli/DTT sample buffer. The IP'ed protein and 10% of the input lysate were subject to SDS-PAGE analysis as above, and protein was detected by staining for the protein tag (e.g. GFP).

Dot blot validation of candidate autoantigens

Individual clones were made by synthesizing the peptide-encoding insert as a single, long DNA oligo (IDT, Ultramer™) that was PCR amplified and then cloned into T7FNS2 in the same way as described for the library. Clones were sequence verified and titered. 2 μl of each clone, after normalizing for titer, was spotted directly onto a nitrocellulose membrane and allowed to dry for 30 minutes. Membranes were blocked with 5% milk, TBST for 1 hour at room temperature, and then stained overnight at 4°C with 1 μg/ml of CSF antibody diluted in a solution containing a 1:1 mix of 5% milk, TBST and T7 10-3b-FLAG phage lysate. Human antibody from CSF was then detected with 1:3,000 peroxidase-conjugated goat affinity purified anti-Human IGG (whole molecule) secondary antibody (MP Biomedicals, 55252) in 2.5% milk, TBST. Quantification was

performed by scanning developed films and analyzing the .tiff file with Image J software.

SUPPLEMENTAL REFERENCES

33. LeProust, E.M. et al. Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res* **38**, 2522-2540 (2010).